Naive Bayes from Scratch:

Bayes Theorem:

★ $$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probability of event A happening when event B is true

★ In the terms of y and X

$$P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)}$$

(x is feature vector $X = [x_1, x_2, x_3, x_4 \ldots x_n]$)

Assuming that all features are mutually independent

$$P(y|X) = \frac{P(x_1|y) \cdot P(x_2|y) \cdots P(x_n|y) \cdot P(y)}{P(X)}$$

Select class with highest probability:

$$y = \arg\max_y P(y|X) = \arg\max_y \frac{P(x_1|y)\ldots P(x_n|y) \cdot P(y)}{P(X)}$$

★ Since we are only interested in y, we don't need $P(x)$

$$y = P(x_1|y) \cdot P(x_2|y) \ldots P(x_n|y)$$

★ Since each of these values individually lie b/w 0 and 1, the values can get very small and we can encounter overflow error.

$\therefore$ we take log of each value

$$y = \log(P(x_1|y)) + \log(P(x_2|y)) + \log(P(x_3|y)) + \dots \log(P(x_n|y))$$

Class Conditional Probability

$P(x_i|y)$

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot \exp\left(-\left(\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)\right)$$

σ = variance of y

μ = mean value